

This is a repository copy of *Source apportionment advances using polar plots of bivariate correlation and regression statistics*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/105308/>

Version: Accepted Version

Article:

Grange, Stuart Kenneth, Lewis, Alastair orcid.org/0000-0002-4075-3651 and Carslaw, David Carlin orcid.org/0000-0003-0991-950X (2016) Source apportionment advances using polar plots of bivariate correlation and regression statistics. *Atmospheric Environment*. pp. 128-134. ISSN 1352-2310

<https://doi.org/10.1016/j.atmosenv.2016.09.016>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Source apportionment advances using polar plots of bivariate correlation and regression statistics

Stuart K. Grange^{a,*}, Alastair C. Lewis^{a,b}, David C. Carslaw^{a,c}

^a*Wolfson Atmospheric Chemistry Laboratory, University of York, York, YO10 5DD, United Kingdom*

^b*National Centre for Atmospheric Science, University of York, Heslington, York, YO10 5DD, United Kingdom*

^c*Ricardo Energy & Environment, Harwell, Oxfordshire, OX11 0QR, United Kingdom*

Abstract

This paper outlines the development of enhanced bivariate polar plots that allow the concentrations of two pollutants to be compared using pair-wise statistics for exploring the sources of atmospheric pollutants. The new method combines bivariate polar plots, which provide source characteristic information, with pair-wise statistics that provide information on how two pollutants are related to one another. The pair-wise statistics implemented include weighted Pearson correlation and slope from two linear regression methods. The development uses a Gaussian kernel to locally weight the statistical calculations on a wind speed-direction surface together with variable-scaling. Example applications of the enhanced polar plots are presented by using routine air quality data for two monitoring sites in London, United Kingdom for a single year (2013). The London examples demonstrate that the combination of bivariate polar plots, correlation, and regression techniques can offer considerable insight into air pollution source characteristics, which would be missed if only scatter plots and mean polar plots were used for analysis. Specifically, using correlation and slopes as pair-wise statistics, long-range transport processes were isolated and black carbon (BC) contributions to PM_{2.5} for a kerbside monitoring location were quantified. Wider applications and future advancements are also discussed.

Keywords:

Air quality, Relationships, Robust regression, Particulate matter, Black carbon

1. Introduction

Determining how variables are related to one-another is a key component of data analysis and statistics. Within the atmospheric sciences, exploring the relationships between chemical constituents and meteorological parameters is extremely common and the techniques for comparing, correlating, and determining relationships are very diverse. Analysis involving the correlation of two pollutants can often be insightful because it can lead to the identification of emission source characteristics, as can investigation into ratios or slopes from regression analysis between two pollutants (Statheropoulos et al., 1998). Within atmospheric disciplines, data analysis can also benefit from being able to integrate wind behaviour (Elminir, 2005). The use of wind speed and direction can be informative because it often leads to the suggestion of source locations and source characteristics, such as height of emission above the surface (Henry et al., 2002; Westmoreland et al., 2007).

Exploration of relationships among variables can be achieved with many different methods that can range from the simple to numerically complex. However, a technique that is used very widely is the simple x - y scatter plot (Bentley, 2004). Scatter plots are useful because they allow for the visualisation of variables and model fitting can be evaluated quickly and simply with visual feedback. Regression techniques, most commonly ordinary least-squared regression, are often employed to formally quantify how x and y are related. The use of least-squared regression is however technically questionable in many cases, and despite a large collection of alternative techniques available, its use remains a persistent feature of air quality data analysis. The use of simple scatter plots is usually carried out with entire datasets or with simple or superficial filtering and therefore have potential to hide some discrete relationships which are present in the global data if they do not conform to the mean rate of change (Cade and Noon, 2003).

Slopes from regression models relating two pollutants to one another are often used in applications that use monitoring data such as emission inventories and pollutant models.

*Corresponding author

Email addresses: `skg511@york.ac.uk` (Stuart K. Grange), `david.carslaw@york.ac.uk` (David C. Carslaw)

When measurements are not available, slopes for the unknown pollutants are often substituted from the literature, short-term monitoring, or data collected at a near-by location. However, the use of simple and static ratios is likely to be deficient in many situations because they can be expected to be highly dependent on source (Manoli et al., 2002). To differentiate sources in air quality data, techniques other than simple scatter plots often need to be used.

A common method for source characterisation is the use of bivariate polar plots (Carslaw et al., 2006; Westmoreland et al., 2007; Carslaw and Beevers, 2013; Uria Tellaetxe and Carslaw, 2014). Polar plots are typically used to visualise and explore mean pollutant concentrations for single species based on wind speed and wind direction. In the atmospheric sciences, it is intuitive to plot wind direction (from 0 to 360° clockwise from north) on the angular ‘axis’ and wind speed to be used for the radial scale. Aggregation functions other than the arithmetic mean can be used and different variables apart from wind speed can be used for the radial scale. For example, atmospheric temperature or stability are often useful variables to use. The main attribute for the choice of radial-axis variable is that it helps to differentiate between different source characteristics in some way due to different source types responding differently to values of the angular scale. Despite the range of potential options, wind speed is widely used to help discriminate different source types and is particularly useful when used together with wind direction and the concentration of a species (Harrison et al., 2001; Kassomenos et al., 2012).

This type of polar plot analysis has, in part, become wide-spread due to the open-source `polarPlot` function available in the *openair* R package (Carslaw and Ropkins, 2012; R Core Team, 2016). Other similar techniques such as non-parametric wind regression have also shown their ability to determine source locations for various pollutants by using polar plots (Henry et al., 2002, 2009; Donnelly et al., 2011).

1.1. Objectives

Combining correlation and regression techniques with those that provide information on source apportionment potentially offers considerably more insight into air pollution sources. The use of wind behaviour has the potential to evaluate correlation and slopes based on

source locations and therefore different processes. It is common for emission inventories to use ratios for pollutants when they are not measured or when high quality data is lacking. It is hypothesised that the combination of correlation, regression, and polar plots could lead to significant additions to data analysis by understanding how different pollutants are related to one another depending on source.

In this paper, the combination of bivariate polar plots approaches with correlation and regression techniques is considered for comparing two pollutants. This combination of methods is then used to aid the interpretation of air quality data. The primary objectives of this paper are as follows. First, to develop methods to combine bivariate polar plot techniques with correlation and a range of linear regression approaches. Second, apply the methods to commonly available measurements of air pollutants to demonstrate the new insights made possible by these techniques. Third, to consider the wider potential uses of the approaches in air quality science. The software developed has been released with an open-source licence and can be found in the *polarplotr* R package ([Carslaw and Grange, 2016](#)).

2. Methods

2.1. Function development

2.1.1. Kernel weighting and scaling

The plotting mechanism for polar plots when using wind direction as the polar axis generally involves first aggregating a time-series into wind speed and direction intervals (or ‘bins’). The specific intervals and numbers of the bins can be altered for a particular application, but all combinations of the two types of bins are summarised by an aggregation function such as the mean or maximum. In the *openair* `polarPlot` function, a smoothed surface is fitted to these binned summaries using a generalised additive model (GAM) to create a continuous surface which can be plotted with polar coordinates. Further details of the approach can be found in [Carslaw and Beevers \(2013\)](#) and [Uria Tellaetxe and Carslaw \(2014\)](#).

When applying a simple aggregation function, the number of observations in a time-series which compose a discrete wind speed and direction bin is not critical for the calculation or the

visual presentation of the surface, except at the edges of the plot where there are (usually) few observations. However, when calculating correlations or relationships between two variables, it becomes important to consider the minimal number of observations which would create a valid summary. If there are too few observations for a particular bin and a statistic such as the correlation or slope is calculated between a pair of variables, it is likely that unreliable summaries will be generated due to large variations between neighbouring bins. To overcome this limitation, for each wind speed and direction bin, the entire time-series was evaluated but observations were *weighted* by their proximity to a wind speed and direction bin *i.e.*, wind speed or direction values further from the bin centre are weighted less than those closer to the centre of the bin. Like previous works such as Henry et al. (2002, 2009), a weighting kernel was used to create weighting variables.

The weighting kernel used was the Gaussian kernel (Equation 1). The Gaussian kernel has infinite tails and therefore all input bins are given a non-zero weighting, but observations furthest from the bin being analysed have very small weights associated with them. The Gaussian kernel was used for weighting both wind speed and direction because it is considered more utilitarian than many other kernels such as the Epanechnikov kernel which have finite bounds and therefore at times, will give observations weights of zero which can cause ambiguity issues.

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (1)$$

To ensure the weighing variable was appropriate for the particular wind speed and direction application, the input wind speed and direction variables required scaling. The scaling process used was simple; the wind variables were multiplied by an integer to increase their bounds and therefore influence within the weighting kernel. The variables were also normalised to ensure that all observations had values between zero and one. This normalisation step is not strictly necessary when the Gaussian kernel is used, but is needed for some other kernels and ensures the output of process always had a known range.

If the weighting operated too locally, the inherently variable nature of wind behaviour

109 was represented in the plotted surface as noise. Conversely, if weighting was extended too far,
 110 isolated areas of ‘real’ peaks were obscured due to over-smoothing. It is difficult to determine
 111 an optimal set of scaling values for wind speed and direction for every application, therefore
 112 a series of heuristic simulations were performed to determine the ideal integer scaling values.

113 It was found that within a central range the final output was rather insensitive to the
 114 scaling values. One reason for this relative insensitivity will be due to the inherent random
 115 variability of concentrations as a function of either wind speed or wind direction due to
 116 atmospheric turbulence. This indicates that within a central band of values, the scaling
 117 process is not particularly influential. It is possible for other applications these scaling
 118 magnitudes will have to be tuned and therefore the defaults can be altered by the user.
 119 An example of the scaling defaults used in the `polarPlot` function are shown in Figure 1.
 120 Figure 1 allows visualisation of the Gaussian weighting kernel for both the wind speed and
 121 direction variables as well as the extent of the default scaling procedure for a single bin for
 122 4.8 m s^{-1} and 230 degrees.

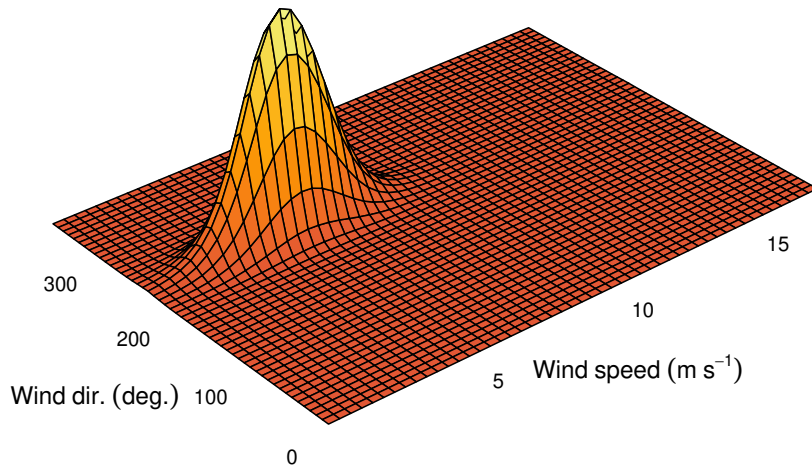


Figure 1: Three-dimensional surface of weights for a single wind speed and direction bin (4.8 m s^{-1} and 230 degrees respectively). The surface is normalised and therefore intensity units are not informative.

123 After the appropriate weights have been calculated, the calculation of any pair-wise
 124 statistic that allows for weighting could be calculated between two pollutants. The first
 125 methods implemented were the Pearson correlation coefficient and two linear regression
 126 methods. Using these two groups of techniques allowed for the investigation of the correlation

between two pollutants and the investigation of the slope between pollutants, but with the inclusion of wind speed and direction.

2.1.2. Correlation

Correlation is a measure of how well two (or more) variables are associated to one-another. Correlation is a useful measure for air pollutants because pollutants which demonstrate high levels of correlation are often emitted from the same source, or undergo similar chemical and physical transformations in the atmosphere. For use in polar plots, the correlation statistic implemented was the weighted Pearson correlation coefficient (r) (Davison and Hinkley, 1997; Canty and Ripley, 2016).

2.1.3. Regression

Regression is a very common statistical technique and is often used to describe and investigate relationships among variables (Kariya and Kurata, 2004). Regression is a large topic and only the linear regression techniques considered for the polar plot function will be discussed. Of particular interest is the estimate of the slope from a linear regression between two species. The slope will often reveal useful information concerning source characteristics, for example, the amount of PM₁₀ that is in the fine fraction (PM_{2.5}), or the ratios of combustion products such as CO and NO_x which can be compared with emission inventory estimates.

The first regression technique implemented was weighted least-squares linear regression. This is very similar to ordinary least-squares linear regression, but the weighted sum of squares are minimised which has the effect of creating a model which preferentially represents a local area of the input data rather than the entire set. Because of the common presence of outliers in air pollution time-series measurements, other regression methods such as robust regression can offer advantages over the least-squares regression for use in the enhanced polar plots.

Robust regression extends least-squares regression techniques in attempting to better handle situations where the parametric assumptions of the least-squares regression method

are violated. These violations are usually involved with the presence of outliers and heteroscedasticity (non-equal variances). Primarily, the power of robust regression lies in the resistance to the influence of outliers. Robust regression achieves this by substituting the least-squares estimator for a more robust estimator (Yohai, 1987). There are many types of robust estimators, but they all operate by first classing observations as outliers or not-outliers and then reducing the influence of the outliers on the regression model (Huber, 1973). The procedures for calculating robust estimators are iterative and more computationally demanding when compared to the calculation of the least-squares estimator. This is noticeable to a user of the `polarPlot` function because additional run-time is needed when the robust regression techniques are used. The robust regression functions were supplied by the *MASS* package (Venables and Ripley, 2002) and the estimator used was the M-estimator because this estimator allows the use of weights.

2.2. Data

Data analysis was conducted on hourly air quality monitoring data for two sites included in the United Kingdom’s Automatic Urban and Rural (AURN) Network. The two sites were London Marylebone Road and London North Kensington (Table 1 and Figure 2). Monitoring data for 2013 were downloaded using the `openair importAURN` function. Both monitoring sites measure a large complement of chemical and particulate species and achieve high data capture rates. The particulate matter measurements were focused on for polar plot analysis and PM_{10} and $PM_{2.5}$ at London Marylebone Road and London North Kensington are monitored by TEOM-FDMS (Tapered Element Oscillating Microbalance-Filter Dynamics Measurement System) instruments. This enhanced method is not as susceptible to removing volatile and semi-volatile components in the monitored air-stream as standard heated TEOMs (Allen et al., 1997; Green et al., 2009). Hourly black carbon (BC) data were also used and these data were sourced directly from the AURN monitoring database after personal communication with Ricardo Energy & Environment. More detailed site and instrument details can be found see at <https://uk-air.defra.gov.uk/>.

Meteorological data for 2013 from London Heathrow (a major airport) in western London

Table 1: Details of locations of air quality and meteorological monitoring sites in London providing data for this study.

Site name	Latitude	Longitude	Elevation	Site type
London North Kensington	51.5211	-0.2134	5	Urban background
London Marylebone Road	51.5225	-0.1546	35	Urban traffic
London Heathrow	51.4780	-0.4610	25.3	Meteorological only



Figure 2: Locations of air quality and meteorological monitoring sites in London providing data for this study. The map’s internal polygons show London’s Boroughs, the City of London, and the River Thames.

were used to represent regional conditions for the two air quality monitoring sites. Hourly data from the London Heathrow site were obtained from the NOAA Integrated Surface Database (ISD) and access was gained with the *worldmet* R package (NOAA, 2016; Carslaw, 2016). The data from Heathrow Airport were used in preference to other local surface measurements, which tend to be strongly influenced by local buildings.

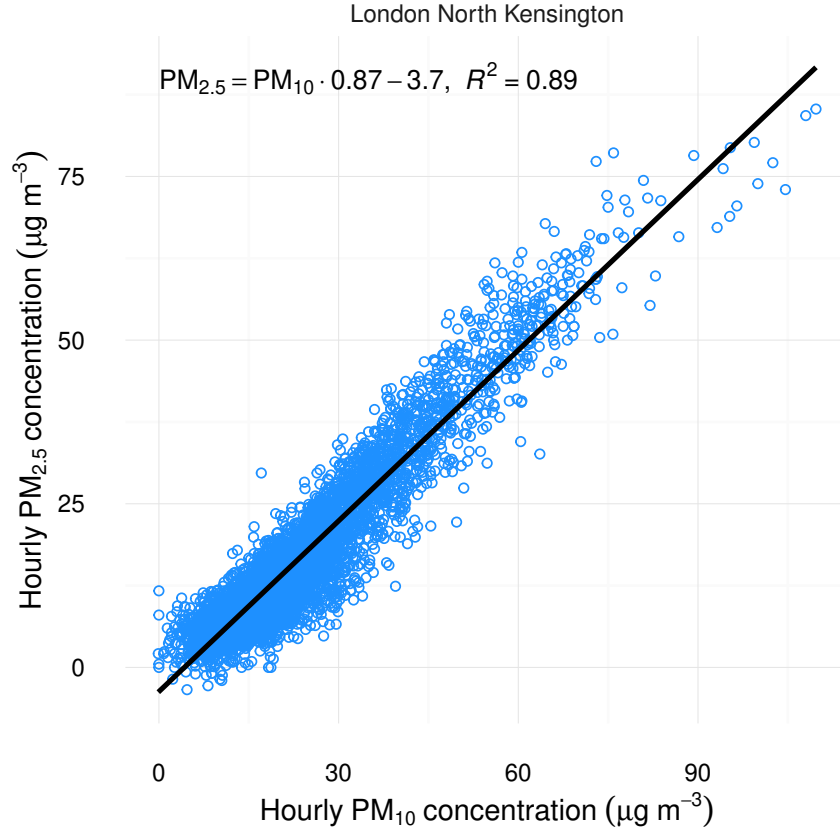


Figure 3: Simple x - y scatter plot of $PM_{2.5}$ and PM_{10} for 2013 at London North Kensington. Fitted line and equation represents the ordinary least-squared regression model.

3. Results & discussion

3.1. London North Kensington PM_{10} and $PM_{2.5}$

London North Kensington is an urban background site (Table 1 and Figure 2) and it is expected that a wide range of sources will contribute particle concentrations, including both local (London) and long-range (continental Europe) sources. A scatter plot of $PM_{2.5}$ and PM_{10} shows that the two particle size fractions showed a good degree of correlation during 2013 (Figure 3). From Figure 3 alone there is no obvious indication that different source types contribute to the overall scatter of points. The mean ratio between $PM_{2.5}$ and PM_{10} was 0.87, as determined by the ordinary least-squares linear regression model and it explained 89% of the variation (Figure 3).

197 The usual use of polar plots, by calculating the mean concentration for wind speed and
 198 directions bins, show that there were multiple sources of PM_{10} and $\text{PM}_{2.5}$ at London
 199 North Kensington in 2013 (Figure 4a and Figure 4b). Figure 4 suggests that locally-sourced
 200 particulate matter were present, as potentially indicated by the elevated concentrations at
 201 low wind speeds, but the highest concentrations were experienced with easterly winds when
 202 wind speeds were high ($\approx 10 \text{ m s}^{-1}$). By contrast, NO_x , a pollutant which is dominated
 203 by local (London) emissions, showed that only when wind speeds were low, were elevated
 204 concentrations experienced due to a lack of pollutant dispersion (Figure 4c). However, when
 205 the $\text{PM}_{2.5}$ and PM_{10} data are plotted with a correlation statistic binned by wind speed and
 206 direction, the situation is more revealing than the scatter plot and mean polar plots would
 207 suggest alone (Figure 5).

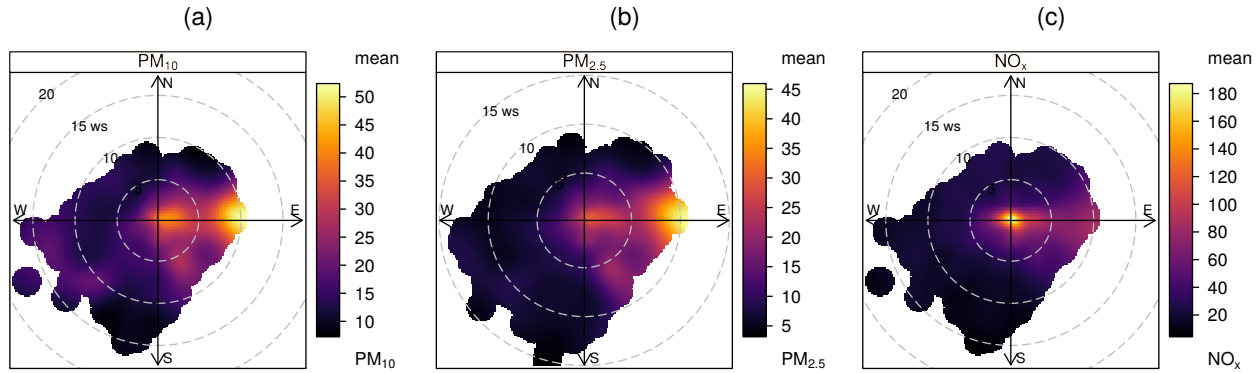


Figure 4: Polar plots of mean concentrations of PM_{10} (a), $\text{PM}_{2.5}$ (b), and NO_x (c) for 2013 at London North Kensington.

208 The correlation polar plot of $\text{PM}_{2.5}$ and PM_{10} demonstrates that during easterly winds,
 209 the London North Kensington $\text{PM}_{2.5}$ and PM_{10} concentrations were very highly correlated
 210 with $r \approx 0.9$ (Figure 5). The zone of high correlation is interpreted to be due to long-range
 211 transport which is characterised by the majority of PM_{10} being made up of $\text{PM}_{2.5}$. In London,
 212 and most areas of the UK, long-range transport is most important under easterly conditions
 213 where air-masses originate from continental Europe (Buchanan et al., 2002; Abdalmogith and
 214 Harrison, 2005; Liu and Harrison, 2011). Under these conditions the concentrations of fine
 215 particulate sulphate and nitrate can dominate absolute particle concentrations. The surface of

Figure 5 is also smooth and covers a wide range of wind speed and directions which indicates a general, and large-scale process which is being appropriately smoothed and represented by the weighting procedure (Section 2.1). Other monitoring locations, including London Marylebone Road that also measure $\text{PM}_{2.5}$ and PM_{10} showed similar easterly behaviour (not shown).

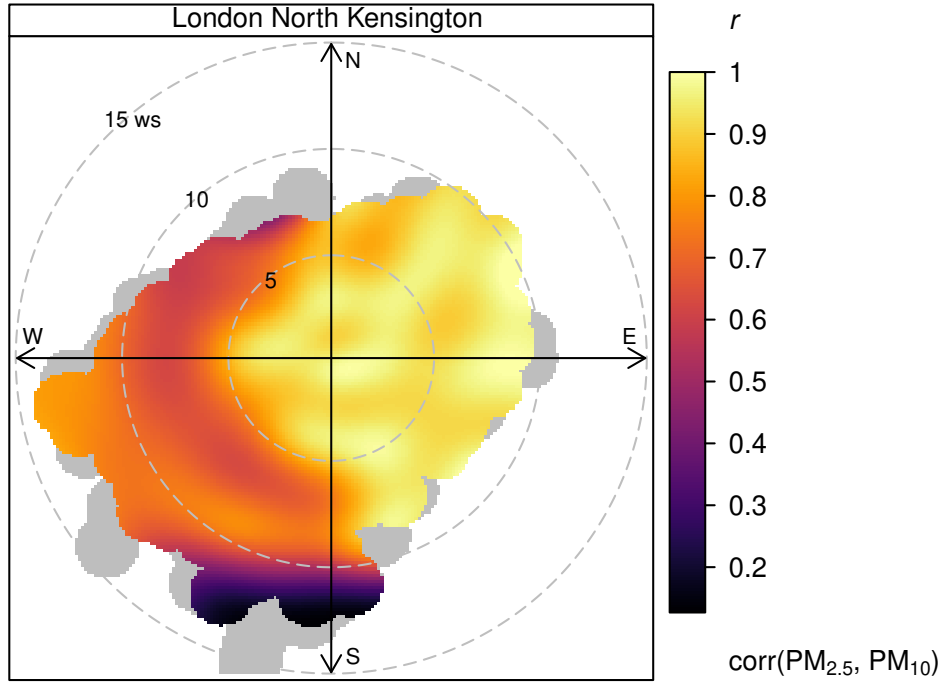


Figure 5: Polar plot of the correlation between $\text{PM}_{2.5}$ and PM_{10} for 2013 at London North Kensington.

Previous studies such as Querol et al. (2004); Charron and Harrison (2005); Harrison et al. (2001); Liu and Harrison (2011) have reported high $\text{PM}_{2.5}$ – PM_{10} ratios for European sourced particulate matter in the UK and the correlation presented in Figure 5 is consistent with these past works which reported high $\text{PM}_{2.5}$ – PM_{10} ratios. When HYSPLIT (Stein et al., 2015) back-trajectories for 2013 were clustered and joined to coincident pollutant observations, the cluster representing air-masses from Europe also had the highest $\text{PM}_{2.5}$ – PM_{10} ratio of all clusters, consistent with the conclusions inferred from Figure 5.

The polar plot of the slope between $\text{PM}_{2.5}$ and PM_{10} at London North Kensington demonstrates a similar surface pattern as the correlation polar plot (Figure 6). The long-range sourced particulate from the east was indeed primarily composed of $\text{PM}_{2.5}$, as shown

231 by a $\text{PM}_{2.5}$ to PM_{10} slope of about 90 %. For other wind directions, coarser particulate
 232 matter was a more important contributor to PM_{10} and the $\text{PM}_{2.5}$ contributions drop to
 233 approximately 30 % (Figure 6). This reduction of $\text{PM}_{2.5}$ to PM_{10} slope was most likely caused
 234 the local process of mechanical resuspension. Even though the scatter plot of $\text{PM}_{2.5}$ and
 235 PM_{10} (Figure 3) does not indicate different source influences, it is clear from Figure 6 in
 236 particular that there are at least two major source types affecting particulate concentrations
 237 at the London North Kensington site. It should be noted that a careful wind speed, wind
 238 direction subset of the data shown in Figure 3 does confirm the behaviour seen in Figure 6
 239 with a much lower $\text{PM}_{2.5}$ to PM_{10} slope for south-westerly winds above 5 m s^{-1} .

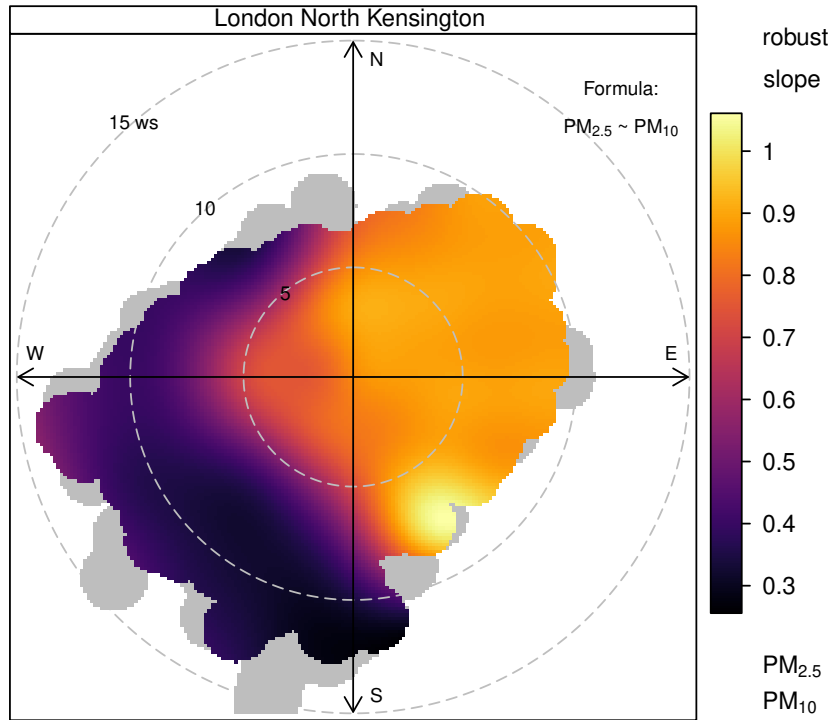


Figure 6: Polar plot of the robust slope between $\text{PM}_{2.5}$ and PM_{10} for 2013 at London North Kensington.

240 3.2. London Marylebone $\text{PM}_{2.5}$ and BC

241 Unlike PM_{10} and $\text{PM}_{2.5}$ at London North Kensington, the London Marylebone Road BC
 242 and $\text{PM}_{2.5}$ correlation was poor in 2013, as shown in Figure 7. Although BC exists primarily
 243 within the fine particle fraction (Petzold et al., 1997; Viidanoja et al., 2002) and would be

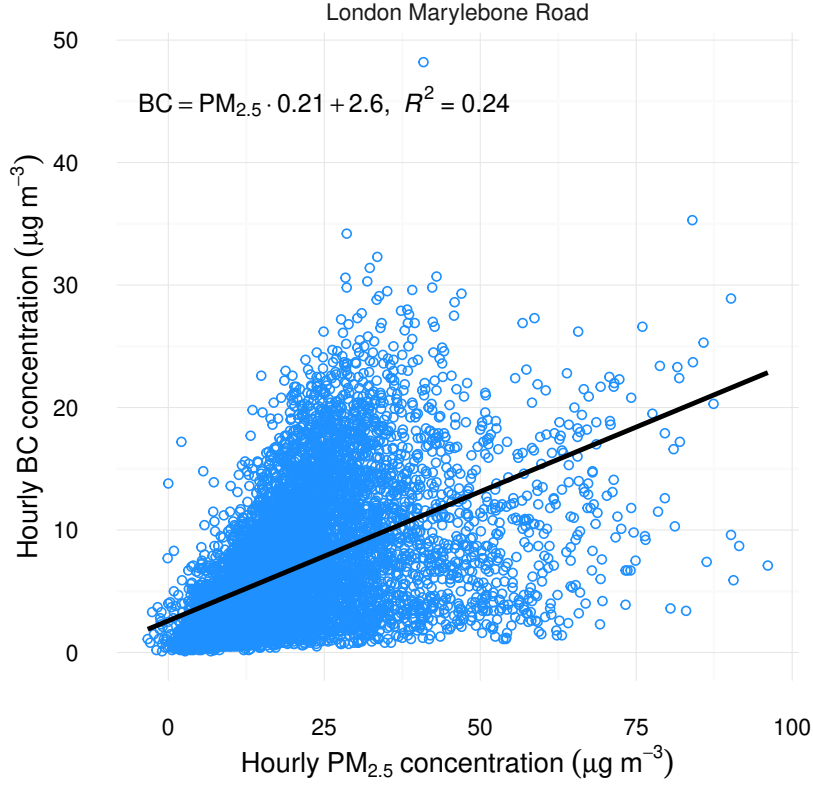


Figure 7: Simple x - y scatter plot of BC and $PM_{2.5}$ for 2013 at London Marylebone Road. Fitted line and equation represents the ordinary least-squared regression model.

expected to be an important component of $PM_{2.5}$ at a traffic-dominated location like London Marylebone Road, $PM_{2.5}$ also has a diverse number of other sources including secondary inorganic aerosol (Querol et al., 2004). Therefore, at times, BC will be a major contributor to $PM_{2.5}$ while at others it will be a minor component depending on the strength of the various sources. Using a scatter plot to investigate this relationship is not immediately useful because the two variables do not follow a mean rate of change. Therefore, fitting a simple linear regression line to these data is not informative (Figure 7).

The robust regression slope of BC and $PM_{2.5}$ binned by wind speed and direction at London Marylebone Road demonstrated patterns that were not observed by the simple scatter plot alone (Figure 8a). Figure 8a shows that the ratio between BC and $PM_{2.5}$ was highly dependent on wind direction. Winds from the south and west at London Marylebone

255 Road had a higher ratio of BC with $\approx 50\%$ of $\text{PM}_{2.5}$ being composed of BC. BC- $\text{PM}_{2.5}$
 256 ratios are sparsely reported, however London Marylebone Road's ratio is consistent with
 257 what [Ruellan and Cachier \(2001\)](#) reported for a traffic-dominated monitoring location in
 258 Paris (Porte d'Auteuil) with ratios of $43 \pm 20\%$. When winds were from the north and
 259 westerly directions, the BC- $\text{PM}_{2.5}$ ratio was lower, usually under 20% . Additionally, winds
 260 from the north were nearly completely free of BC particulate matter (Figure 8a).

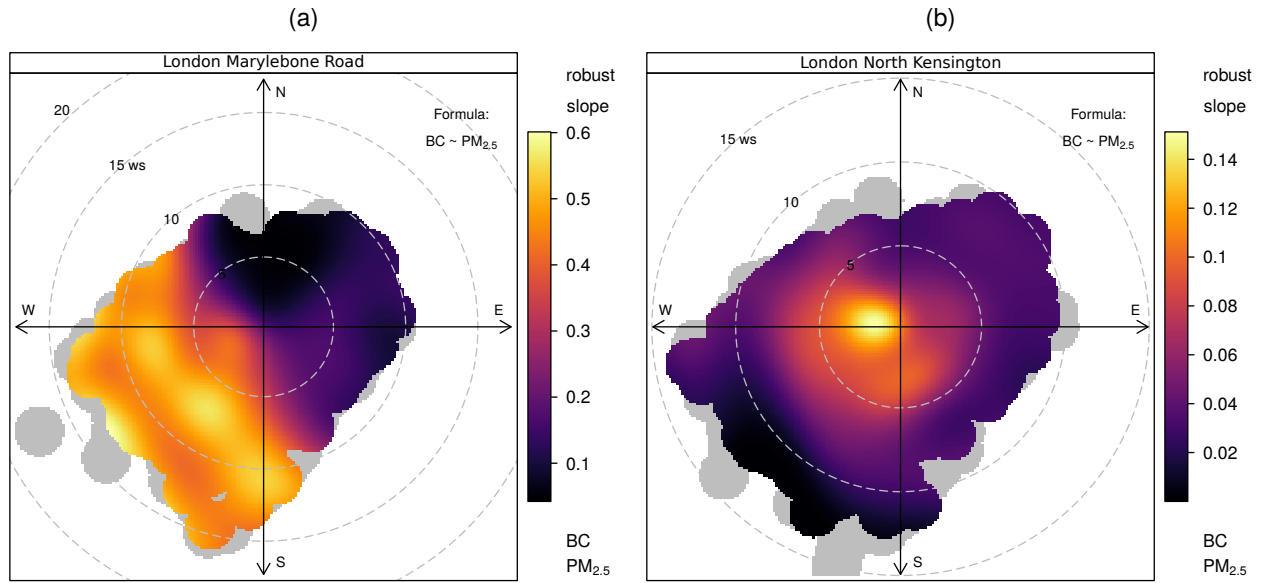


Figure 8: Polar plot of the robust slope between BC and $\text{PM}_{2.5}$ at London Marylebone Road (a) and London North Kensington (b).

261 The wind direction dependencies inferred from the polar plot are somewhat counter-
 262 intuitive given that the London Marylebone Road monitoring site is located one metre from
 263 the kerb on the south-side of an arterial road. However, the site is also within a significant
 264 street-canyon with a width of 40 m and a height of 41 m which is likely to lead to complex
 265 recirculation patterns at a range of wind speeds ([Charron and Harrison, 2005](#); [Giorio et al., 2015](#)).
 266 Based on this evidence, accumulation of pollutants on the buildings' lee-side (south)
 267 is an important process to consider at London Marylebone Road when interpreting source
 268 processes.

269 London North Kensington also measures BC and $\text{PM}_{2.5}$ and the slope of these two
 270 pollutants binned by wind speed is rather different compared with London Marylebone Road

(Figure 8b). London North Kensington is an urban background site and lacks the large traffic source being in immediate proximity which London Marylebone Road experiences. Therefore, BC was a much smaller component of $\text{PM}_{2.5}$. In 2013, London North Kensington had a maximum contribution of $\approx 15\%$ of BC to $\text{PM}_{2.5}$ (Figure 8b). However, this maximum contribution only occurred when wind speeds were low and suggests that this contribution is reached only when local traffic emissions influence the monitoring site.

Based on these results for the two monitoring sites, the clear and consistent BC- $\text{PM}_{2.5}$ ratio at London Marylebone Road of around 50% shown in Figure 8a in the south-west quadrant can be interpreted as a contribution dominated by local traffic sources. The lower ratio of between 10–20% mostly to the east is dominated by regional source contributions where the concentration of $\text{PM}_{2.5}$ is relatively high but where air masses contain very little BC.

3.3. Future directions

The examples presented for a single year of data for two air quality monitoring sites in London were the first steps for enhancing polar plots to include the functionality of pair-wise statistics. The enhancements were able to substantially improve the information content available from routinely monitored air pollutants where simple scatter plots and ‘standard’ polar plots gave no suggestion of the processes subsequently illuminated by the correlation/slope polar plots.

The examples reported were for a few commonly measured species. However, it is expected that the use of polar plots using pair-wise statistics for multi-species data such as metal or VOC concentrations could be highly informative. Measurement of large numbers of metals and other species at higher time resolutions (hourly) is becoming more common. A ‘correlation matrix of robust slope polar plots’ would potentially reveal more detailed information on common source origins.

The use of other statistics is another valuable future direction such as non-parametric measures of correlation such as Spearman. Other regression techniques such as quantile regression (Koenker and Bassett, 1978) could be implemented to provide slope information

299 across a range of quantile levels, potentially providing more comprehensive information on
300 the relationship between two pollutants and give further options when determining pollutant
301 sources. The main advantage of quantile regression is likely to be related to resolving two
302 or more sources that overlap and where there is not a single dominant slope caused by
303 one source. In this case, considering the full distribution of slope values may help better
304 resolve competing source contributions. Finally, the weighted statistics approach for paired
305 statistics could usefully be extended to model evaluation where two sets of data are compared
306 (observed and modelled). In this case, enhanced polar plot analyses could provide valuable
307 information concerning where model agreement is good or poor and indicate more clearly the
308 conditions under which model performance is acceptable and provide enhanced information
309 on where model performance is poor.

310 4. Conclusions

311 This paper outlined the development of enhanced bivariate polar plots to include pair-wise
312 statistics to be used in the atmospheric sciences. Two groups of statistical techniques were
313 implemented: correlation and regression. The new development brings together commonly
314 used pair-wise statistics and relationships with wind speed and direction, which provides
315 enhanced information on pollutant sources beyond currently used techniques.

316 Using a single year of data, in a single city, for routinely monitored pollutants demonstrated
317 that the enhanced polar plots were capable of determining relationships and processes that
318 were not suggested by simple scatter plots and the use of mean polar plots alone. Here we
319 have reported that traffic dominated $\text{PM}_{2.5}$ is composed of 50 % BC at a London monitoring
320 site. This is an important observation and ratios between other pollutants such as elemental
321 carbon and organic carbon (EC and OC) is an obvious future application for the enhanced
322 polar plots.

323 It is expected in the future that enhanced polar plots will be widely used for the
324 investigation of ratios for pairs of pollutants and further extended to be a valuable tool for
325 teasing apart pollutant sources and processes.

Acknowledgements

This work was supported by Anthony Wild with the provision of the Wild Fund Scholarship. This work was also partially funded by the 2016 Natural Environment Research Council (NERC) air quality studentships programme.

Highlights

- Bivariate polar plots are a common method for exploring pollutant sources.
- Polar plots were enhanced with the addition of pair-wise statistics.
- Usage examples of the enhanced polar plots are given for two London monitoring sites.
- Processes were illuminated that were not detected by other plotting methods.
- Potential future applications and extensions are discussed for bivariate polar plots.

References

- Abdalmogith, S. S., Harrison, R. M., 2005. The use of trajectory cluster analysis to examine the long-range transport of secondary inorganic aerosol in the UK. *Atmospheric Environment* 39 (35), 6686–6695.
URL <http://www.sciencedirect.com/science/article/pii/S1352231005006825>
- Allen, G., Sioutas, C., Koutrakis, P., Reiss, R., Lurmann, F. W., Roberts, P. T., 1997. Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas. *Journal of the Air & Waste Management Association* 47 (6), 682–689.
- Bentley, S., 2004. Graphical techniques for constraining estimates of aerosol emissions from motor vehicles using air monitoring network data. *Atmospheric Environment* 38 (10), 1491–1500.
URL <http://www.sciencedirect.com/science/article/pii/S1352231003010574>
- Buchanan, C., Beverland, I., Heal, M., 2002. The influence of weather-type and long-range transport on airborne particle concentrations in Edinburgh, UK. *Atmospheric Environment* 36 (34), 5343–5354.
URL <http://www.sciencedirect.com/science/article/pii/S1352231002005794>
- Cade, B. S., Noon, B. R., 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1 (8), 412–420.
URL [http://dx.doi.org/10.1890/1540-9295\(2003\)001\[0412:AGITQR\]2.0.CO;2](http://dx.doi.org/10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2)
- Canty, A., Ripley, B. D., 2016. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-18.

353 Carslaw, D., 2016. worldmet: Import Surface Meteorological Data from NOAA Integrated Surface Database
 354 (ISD). R package version 0.6.
 355 URL <http://github.com/davidcarslaw/worldmet>
 356 Carslaw, D., Grange, S., 2016. polarplotr: Functions to plot polar-plots. R package.
 357 URL <https://github.com/davidcarslaw/polarplotr>
 358 Carslaw, D. C., Beevers, S. D., 2013. Characterising and understanding emission sources using bivariate
 359 polar plots and k-means clustering. Environmental Modelling & Software 40, 325–329.
 360 URL <http://www.sciencedirect.com/science/article/pii/S136481521200237X>
 361 Carslaw, D. C., Beevers, S. D., Ropkins, K., Bell, M. C., 2006. Detecting and quantifying aircraft and
 362 other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport.
 363 Atmospheric Environment 40 (28), 5424–5434.
 364 URL <http://www.sciencedirect.com/science/article/pii/S1352231006004250>
 365 Carslaw, D. C., Ropkins, K., 2012. *openair* — An R package for air quality data analysis. Environmental
 366 Modelling & Software 27–28 (0), 52–61.
 367 URL <http://www.sciencedirect.com/science/article/pii/S1364815211002064>
 368 Charron, A., Harrison, R. M., 2005. Fine (PM_{2.5}) and Coarse (PM_{2.5–10}) Particulate Matter on A Heavily
 369 Trafficked London Highway: Sources and Processes. Environmental Science & Technology 39 (20), 7768–
 370 7776.
 371 URL <http://dx.doi.org/10.1021/es050462i>
 372 Davison, A. C., Hinkley, D. V., 1997. Bootstrap Methods and Their Applications. Cambridge University
 373 Press, Cambridge, ISBN 0-521-57391-2.
 374 URL <http://statwww.epfl.ch/davison/BMA/>
 375 Donnelly, A., Misstear, B., Broderick, B., 2011. Application of nonparametric regression methods to study
 376 the relationship between NO₂ concentrations and local wind direction and speed at background sites.
 377 Science of The Total Environment 409 (6), 1134–1144.
 378 URL <http://www.sciencedirect.com/science/article/pii/S0048969710012726>
 379 Elminir, H. K., 2005. Dependence of urban air pollutants on meteorology. Science of The Total Environment
 380 350 (1–3), 225–237.
 381 URL <http://www.sciencedirect.com/science/article/pii/S0048969705000732>
 382 Giorio, C., Tapparo, A., Dall’osto, M., Beddows, D. C. S., Esser Gietl, J. K., Healy, R. M., Harrison, R. M.,
 383 2015. Local and Regional Components of Aerosol in a Heavily Trafficked Street Canyon in Central London
 384 Derived from PMF and Cluster Analysis of Single-Particle ATOFMS Spectra. Environmental Science &
 385 Technology 49 (6), 3330–3340.
 386 URL <http://dx.doi.org/10.1021/es506249z>

387 Green, D. C., Fuller, G. W., Baker, T., 2009. Development and validation of the volatile correction model
 388 for PM₁₀—An empirical method for adjusting TEOM measurements for their loss of volatile particulate
 389 matter. *Atmospheric Environment* 43 (13), 2132–2141.
 390 URL <http://www.sciencedirect.com/science/article/pii/S1352231009000557>
 391 Harrison, R. M., Yin, J., Mark, D., Stedman, J., Appleby, R. S., Booker, J., Moorcroft, S., 2001. Studies of
 392 the coarse particle (2.5–10 μ m) component in UK urban atmospheres. *Atmospheric Environment* 35 (21),
 393 3667–3679.
 394 URL <http://www.sciencedirect.com/science/article/pii/S1352231000005264>
 395 Henry, R., Norris, G. A., Vedantham, R., Turner, J. R., 2009. Source Region Identification Using Kernel
 396 Smoothing. *Environmental Science & Technology* 43 (11), 4090–4097.
 397 URL <http://dx.doi.org/10.1021/es8011723>
 398 Henry, R. C., Chang, Y.-S., Spiegelman, C. H., 2002. Locating nearby sources of air pollution by nonparametric
 399 regression of atmospheric concentrations on wind direction. *Atmospheric Environment* 36 (13), 2237–2244.
 400 URL <http://www.sciencedirect.com/science/article/pii/S1352231002001644>
 401 Huber, P. J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*,
 402 799–821.
 403 Kariya, T., Kurata, H., 2004. Generalized least squares. John Wiley & Sons.
 404 Kassomenos, P., Vardoulakis, S., Chaloulakou, A., Grivas, G., Borge, R., Lumbreras, J., 2012. Levels, sources
 405 and seasonality of coarse particles (PM₁₀–PM_{2.5}) in three European capitals — Implications for particulate
 406 pollution control. *Atmospheric Environment* 54 (0), 337–347.
 407 URL <http://www.sciencedirect.com/science/article/pii/S1352231012001665>
 408 Koenker, R., Bassett, Jr, G., 1978. Regression quantiles. *Econometrica: journal of the Econometric Society*,
 409 33–50.
 410 Liu, Y.-J., Harrison, R. M., 2011. Properties of coarse particles in the atmosphere of the United Kingdom.
 411 *Atmospheric Environment* 45 (19), 3267–3276.
 412 URL <http://www.sciencedirect.com/science/article/B6VH3-52G8GT9-1/2/1c4d705b78225fca7c930197cd80c35a>
 413 [1c4d705b78225fca7c930197cd80c35a](http://www.sciencedirect.com/science/article/B6VH3-52G8GT9-1/2/1c4d705b78225fca7c930197cd80c35a)
 414 Manoli, E., Voutsas, D., Samara, C., 2002. Chemical characterization and source identification/apportionment
 415 of fine and coarse air particles in Thessaloniki, Greece. *Atmospheric Environment* 36 (6), 949–961.
 416 URL <http://www.sciencedirect.com/science/article/pii/S1352231001004861>
 417 NOAA, 2016. Integrated Surface Database (ISD).
 418 URL <https://www.ncdc.noaa.gov/isd>
 419 Petzold, A., Kopp, C., Niessner, R., 1997. The dependence of the specific attenuation cross-section on black
 420 carbon mass fraction and particle size. *Atmospheric Environment* 31 (5), 661–672.

421 URL <http://www.sciencedirect.com/science/article/pii/S1352231096002452>
 422 Querol, X., Alastuey, A., Ruiz, C., Artiano, B., Hansson, H., Harrison, R., Buringh, E., ten Brink, H., Lutz,
 423 M., Bruckmann, P., Straehl, P., Schneider, J., 2004. Speciation and origin of PM₁₀ and PM_{2.5} in selected
 424 European cities. *Atmospheric Environment* 38 (38), 6547–6555.
 425 URL <http://www.sciencedirect.com/science/article/pii/S1352231004008143>
 426 R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical
 427 Computing, Vienna, Austria.
 428 URL <https://www.R-project.org/>
 429 Ruellan, S., Cachier, H., 2001. Characterisation of fresh particulate vehicular exhausts near a Paris high flow
 430 road. *Atmospheric Environment* 35 (2), 453–468.
 431 URL <http://www.sciencedirect.com/science/article/pii/S1352231000001102>
 432 Statheropoulos, M., Vassiliadis, N., Pappa, A., 1998. Principal component and canonical correlation analysis
 433 for examining air pollution and meteorological data. *Atmospheric Environment* 32 (6), 1087–1095.
 434 URL <http://www.sciencedirect.com/science/article/pii/S1352231097003774>
 435 Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., Ngan, F., 2015. NOAA’s HYSPLIT
 436 Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society*
 437 96 (12), 2059–2077.
 438 URL <http://dx.doi.org/10.1175/BAMS-D-14-00110.1>
 439 Uria Tellaetxe, I., Carslaw, D. C., 2014. Conditional bivariate probability function for source identification.
 440 *Environmental Modelling & Software* 59 (0), 1–9.
 441 URL <http://www.sciencedirect.com/science/article/pii/S1364815214001339>
 442 Venables, W. N., Ripley, B. D., 2002. *Modern Applied Statistics with S*, 4th Edition. Springer, New York,
 443 ISBN 0-387-95457-0.
 444 URL <http://www.stats.ox.ac.uk/pub/MASS4>
 445 Viidanoja, J., Sillanpää, M., Laakia, J., Kerminen, V.-M., Hillamo, R., Aarnio, P., Koskentalo, T., 2002.
 446 Organic and black carbon in PM_{2.5} and PM₁₀: 1 year of data from an urban site in Helsinki, Finland.
 447 *Atmospheric Environment* 36 (19), 3183–3193.
 448 URL <http://www.sciencedirect.com/science/article/pii/S1352231002002054>
 449 Westmoreland, E. J., Carslaw, N., Carslaw, D. C., Gillah, A., Bates, E., 2007. Analysis of air quality within
 450 a street canyon using statistical and dispersion modelling techniques. *Atmospheric Environment* 41 (39),
 451 9195–9205.
 452 URL <http://www.sciencedirect.com/science/article/pii/S1352231007006863>
 453 Yohai, V. J., 1987. High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals*
 454 *of Statistics* 15 (2), 642–656.

